ESD-TR-84-018

Semiannual Technical Summary

Wideband Integrated Voice/Data Technology

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Defense Advanced Research Projects Agency
under Electronic Systems Division Contract F19628-80-C-0002.

Approved for public release; distribution unlimited.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

WIDEBAND INTEGRATED VOICE/DATA TECHNOLOGY

SEMIANNUAL TECHNICAL SUMMARY REPORT
TO THE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

1 OCTOBER 1983 — 31 MARCH 1984

ISSUED 8 AUGUST 1984

LEXINGTON MASSACHUSETTS

ABSTRACT

This report describes work performed on the Wideband Integrated Voice/Data Technology Program sponsored by the Information Processing Techniques Office of the Defense Advanced Research Projects Agency during the period 1 October 1983 through 31 March 1984.

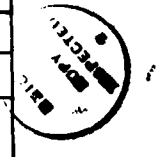

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

INTRODUCTION AND SUMMARY

An important challenge in the design of future military communications networks is to achieve system economy and adaptability through efficient and flexible allocation of common network resources to voice and data users. The major objective of the Wideband Integrated Voice/Data Technology Program is to address this challenge through the development of techniques for integrated voice and data communications in digital packet networks which include wideband common-user satellite links. A major focus of activity in the program has been the establishment of an experimental wideband packet satellite network for realistic testing of a variety of strategies for efficient multiplexing of voice and data users. The program also serves as a focus for the development and testing of techniques for local area packet voice distribution, for speech traffic concentration, and for efficient real-time voice communication in an internetwork environment including local networks of various types connected through a wideband demand-assigned satellite network. Beginning in FY84, an additional focus of the program is on the development of voice-controlled systems, including robust speech-recognition techniques for acoustically noisy environments.

This report covers work in the following areas: development and technology transfer of the generalized LPC (Linear Predictive Coding) Speech-Processing Peripheral (SPP), coordination and execution of multiuser internetwork packet speech experiments using the experimental wideband satellite network (WB SATNET), and development of robust speech recognition technology for voice-controlled systems.

Lincoln prototypes of the SPP are installed and operating at Information Sciences Institute (ISI), Carnegie-Mellon University (CMU), Bolt, Beranek and Newman (BBN), and Adams-Russell (AR). As part of the SPP technology transfer procurement, AR has built and debugged a preprototype breadboard of the SPP. The final prototype PC (Printed Circuit) card and chassis are currently in fabrication, with delivery expected in mid-April. Lincoln's INTEL 8085 software has been modified to accommodate the AR design, and is completely operational.

Lincoln has continued to coordinate the Task Force effort focused on regular multisite WB SATNET operation at 3 Mbps. All seven sites are now operating at 3 Mbps, and the Task Force is working on correcting a few remaining system problems and reliability issues associated with multisite operation. Successful tests have been conducted with three sites running three 64-kbps full-duplex PCM calls in a fully-connected configuration.

The miniconcentrator gateways have continued to operate reliably. An important enhancement has been added to allow more efficient integration of voice and data in WB SATNET streams by using new Multipurpose Packets (MPPs), which can combine speech packets, data packets, and fragments of data packets into a single gateway-to-gateway packet. In addition, the gateways at Lincoln and ISI have been upgraded to allow four network connections, and a variety of upgrades to facilitate automated field use have been added.

Work has been initiated on integrating packet voice onto a standard Ethernet-type local area network. A design for an Ethernet interface which can be inserted into either a packet-voice terminal or a concentrator interface unit has been completed. SUN workstations with Ethernet capability have been installed and integrated in our laboratory, and initial tests with an LPC SPP unit have been conducted.

In the area of voice-controlled systems, we have designed new system structures for recognition of speech in noise, including noise adaptation techniques based on general speech and noise models. Our system designs include a noise-adaptive template-based approach using Dynamic Time Warping (DTW) for time alignment of input and reference, and a robust Hidden Markov Model (HMM) system. We have obtained a noisy-speech data base for the F-16 fighter aircraft environment, and have implemented an LDSP-based (Lincoln Digital Signal Processor) facility for initial experiments. Finally, we have developed level-building software for connected-word recognition, to serve the data-processing function needed to operate in conjunction with the wafer-scale DTW system being developed in our DARPA-sponsored Restructurable VLSI Program.

WIDEBAND INTEGRATED VOICE/DATA TECHNOLOGY

I. SPEECH PROCESSING PERIPHERAL

A. FIELD INSTALLATION

Currently, seven Speech Processing Peripheral (SPP) units are in the field or at Lincoln Laboratory. The locations and delivery dates of the seven units are listed below:

Site	Number of SPPs	Delivery Date
ISI	1	1 February 1983
CMU	1	28 March 1983
AR	1	15 August 1983
BBN	2	1 December 1983
LL	2	

The recent delivery of the BBN units, for applications in multimedia message systems, was carried out at the request of Colonel Bob Baker of DARPA. Adams-Russell is using their unit in debugging their preproduction SPP unit for the technology transfer program.

B. TECHNOLOGY TRANSFER

Adams-Russell currently is executing the six-month prototyping phase (Phase 1) of the SPP technology transfer program. Since the past semiannual report, a preprototype breadboard of the SPP has been built and debugged by AR. This has provided a vehicle for writing and debugging of the final prototype's software as well. The final prototype PC card and chassis are currently in fabrication.

Since substantial electrical changes were made to the Lincoln Laboratory SPP prototype by AR, it was decided to precede the PC board artwork by fabrication of a wirewrap breadboard. This breadboard has been completed by AR and debugged jointly with LL. It has been demonstrated in standalone operation as well as in conjunction with an LL SPP unit. A 3-speaker Diagnostic Rhyme Test (DRT) score of 89.5 (Standard Error 1.16) at 2400 bps was obtained for the AR prototype under clear conditions. In an earlier test, the Lincoln prototype scored 89.1 with a standard error of 1.25. There is no significant difference in performance between the two versions of the SPP. Several minor bugs in the SPP redesign were found in the course of this process and corrected. It is hoped that the breadboard stage will avoid the need for a second pass at the PC board. Currently, the multilayer printed-circuit-board artwork is proceeding along with the design of the prototype chassis.

The Intel 8085-based software interface has been revised to accommodate the Adams-Russell design of the Speech Processing Peripheral. This new version, which is fully operational, includes a set of six diagnostics to be performed whenever a hardware reset is done. An external switch designates slow or fast test mode, and another switch indicates whether or not the board is in loop-back mode. Two of the six diagnostics will be run only if the loop-back switch is set. Three lights plus a test light are used to indicate to the user which diagnostic is being run, so that, if a diagnostic fails, the system will hang, the test light will be on, and the three lights will show the number of the diagnostic that failed. Normally, the system will be powered-up in fast test mode and not in loop-back mode, and a successful diagnostic run will be 'invisible' to the user. If loop-back mode is invoked, the lights for tests 4 and 5 will remain lit for approximately 1/2 second each. In slow test mode, lights for diagnostics 1 through 6 will remain lit for 1/2 second each.

II. WIDEBAND NETWORK EXPERIMENTS AND EXPERIMENT COORDINATION

A. WIDEBAND NETWORK SYSTEM COORDINATION

It has previously been reported that a WB SATNET Task Force was established in March 1983 for the purpose of identifying and correcting the system problems in the network, with the goal of achieving stable and reliable multisite operation at a channel bit rate of 3 Mbps. The Task Force is coordinated by Lincoln, and includes representatives from BBN, ISI and LINKABIT. As the work progressed, it became apparent that the magnitude of the task was quite substantial. The problem solving has tended to require concentrated effort by the whole Task Force at numerous organized on-site work sessions. Also, each round of activity tended to disclose another layer of problems.

This intensive effort has led to significant improvements. There are clear indications that the Task Force goal is indeed within reach. Highlights of the current status of the network are:

(1) All seven of the WB SATNET sites have been upgraded to reflect the results of the debugging efforts of the Task Force, and have been brought to full operation. Due to the subsystem unreliability discussed below, however, the network has not yet been operated with more than five sites on line; a typical network size is three to four sites.

(2) The standard channel bit rate for the network is 3.088 Mbps, which is the nominal design parameter for the system and is higher by a factor of four than had been achieved prior to establishment of the Task Force.

(3) PSAT header and control information is coded at rate 1/2, and speech traffic on the net is typically run at code rate 3/4.

(4) The number of remaining unsolved system problems appears to be small, and it is reasonable to expect that some additional concentrated Task Force effort will eradicate them.

(5) Subsystem unreliability, which has been a major concern, is currently being addressed by equipment upgrades. The ESIs (Earth Station Interfaces) are being replaced by production-quality ESI-As; plans are set for replacement of the PSATs (Packet Satellite Interface Message Processors) by BBN 'Butterfly'-based BSATs; and the newer Western Union earth stations are provided with more reliable equipment, while the older stations are to be upgraded to the same level as the new ones.

Four additional Task Force site visits have been conducted during the present reporting period, in addition to four that were reported in the previous Semiannual Technical Summary. The first of these took place at Lincoln on 14-16 November 1983, with the objectives of correcting certain known problems and further stabilizing system operation at the full channel bit rate of 3.088 Mbps. Substantial progress was made toward these objectives. The next site visit occurred at Fort Monmouth on 5-7 December, where success was achieved in

adding this site to the three already available on the net (Lincoln, ISI and RADC). The next effort was at SRI (SRI International) on 19-20 January 1984, for the purpose of installing an upgraded ESI and restoring the site to operation on the net. The ESI used was the one that had been in place at ISI since WB SATNET implementation first began, and had been made available in mid-December when LINKABIT installed an ESI-A at ISI. The SRI Task Force effort was successful, in that the site was brought up on the net and a number of calls were successfully placed; subsystem failures subsequently occurred, however, keeping SRI from operating again for about two weeks. On 7 February, a landmark event occurred: all five sites then in operating condition (ISI, Lincoln, SRI, RADC and Fort Monmouth) were operated together on the net for an extended period.

The fourth Task Force site visit in this reporting period (and the eighth since inception of the Task Force) was carried out at Fort Huachuca during the period 13-15 February. An ESI-A was installed, and the site was brought into operation on the network. Repeated PCM (Pulse Code Modulation) calls were made among ISI, Lincoln, Fort Monmouth and Fort Huachuca. While this brought the total of active sites to six, no more than four have been operated together since the 7 February experiments.

An attempt was made to restore the seventh site (DCEC) to operation in March by shipping an ESI-A for installation by local personnel. Damage was sustained in shipment, however, and a LINKABIT engineer traveled to DCEC to repair it. The problem was found actually to be in the PSAT, and was repaired; the site was restored to service by the end of March. The largest number of sites operated together on the net is still five.

In response to a DARPA request, the Task Force prepared a White Paper laying out a sequence of recommended actions by sponsor and contractors leading to full realization of an operational WB SATNET, integrated into the DARPA Internet. The White Paper was separated into near-, mid- and long-term sections. It provided a basis for discussion at a formal report to DARPA and DCEC, made by the Task Force in Washington on 6 March. This meeting focused on three central themes: (1) how to bring the network to stable operating condition, (2) how to upgrade to an improved network using BSATs and ESI-Bs, and (3) how to integrate the system into the Internet. Three key items were highlighted with respect to the first theme, namely: use of the PSAT and ESI-A to further stabilize network operation with multiple sites, construction of additional ESI-As to outfit the rest of the net, and completion of certain identified improvements in the earth stations. The second theme involved four key items, namely: development and test of the BSAT and ESI-B, replication of the latter as necessary to outfit the network, replacement of five-meter antennas with seven-meter upgrades, and moving to a higher-power Western Union transponder. The third theme will be the subject of ongoing planning and activity.

B. HOST-LEVEL NETWORK EXPERIMENTS

Tests using traffic generators in the IP/ST (Internet Protocol/Stream Protocol) gateways have shown an apparent upper limit for PSAT handling of relatively short (64 byte) packets in the range of 150 to 160 packets per second. For packets requesting stream service,

attempts to send at higher rates result in loss of offered packets in excess of the limit. For packets requesting datagram service, higher rates result in PSAT crashes. Investigation by BBN showed that the excess stream packets were not seen by the PSAT, suggesting that host module processing in the PSAT is falling behind at high rates, and the end-of-packet flags from the I/O hardware are being missed. Tests with two hosts showed that each could send at the maximum rate without hitting limits in the channel processing capabilities of the PSAT. The observation of this disturbingly low limit in the handling of host packets by the PSAT has provided additional motivation for our development of the multipurpose packet capability described in Section II-C of this report.

In the latter part of this reporting period, it has become routine for there to be three or more operational sites on the WB SATNET, and we are now able to resume voice experiments involving more than two sites. We have been successful in carrying out three-site experiments with three 64-kbps PCM full-duplex conversations, but success in such experiments is not routine, and difficulties are often encountered. The difficulties seem to depend on the total number of sites on the network even though all the sites are not directly involved in the experiment. In some cases, call attempts will consistently succeed if placed in one direction but fail with equal consistency if tried in the reverse direction. Failure usually occurs because the PSAT either does not respond to the gateway's attempt to increase the SATNET stream capacity to handle the call, or it responds so slowly (after many seconds) that the higher level protocols abort the call on timeouts. The probability of success seems to fall as the total committed stream capacity rises, and BBN has expressed the suspicion that the problem may be due to a bug in the PSAT software that fragments WB SATNET datagrams to fit the channel slots that remain after stream allocations have been made. WB SATNET datagrams are not used directly in the voice call setup process, but they do occur indirectly because PSATs that lack ARPANET (ARPA Network) connections send monitoring packets on the channel as WB SATNET datagrams, and such packets are generated as a result of the gateways' attempts to change the stream parameters. The solution to this stream-change problem is an important item on the agenda of the Wideband Task Force.

C. GATEWAY DEVELOPMENTS

Major extensions were made to the Gateway to provide more automated field use, to optimize dispatching of messages to networks, to probe other internet hosts via ICMP (Internet Control Message Protocol) ECHO messages, and to accomodate four networks simultaneously. A prerequisite for the incorporation of these capabilities was the exploitation by the Gateway of the separate I and D space features of the PDP-11 computers. Such usage of separate I and D space was recently made possible in the EPOS operating system through mechanisms provided by ISI personnel.

1. Automated Field Use

A number of extensions were made to the Gateway to automate some of its operations. These extensions are especially desirable for Gateways in the field where individuals unfamiliar with the intricacies of Gateway operation need to support them.

Using new capabilities in the EPOS operating system, as provided by ISI personnel, the Gateway was modified to be automatically invoked when the EPOS operating system is booted on the PDP-11/44. Therefore, the only manual operations needed to invoke the Gateway are to power up the PDP-11/44 and boot the EPOS operating system. The latter step is not needed if switches in the PDP-11/44 are set for automatic booting off the desired device upon power up.

The management of PSAT streams has been automated. When the Gateway achieves communication with the PSAT for the first time after invocation, it clears out any streams that may be have been left from before (whose parameters it does not know) and creates a standard stream of nominal size for later use. After a PSAT down/up cycle, the Gateway re-establishes the streams that were extant before the 'bounce'. A request to establish a point-to-point connection results in automatic enlargement of the standard stream via a request to the PSAT; conversely, the closing of a connection results in the automatic shrinkage of the standard stream. Timeout/retransmission is used with these PSAT requests to assure reliable achievement of the desired results. Additionally, the Gateway supports up to five additional streams that may be needed for measurement experiments.

The Gateway now communicates the current date-time to other Gateways by 'piggybacking' this information in 'group join' messages sent between the Gateways. This permits a newly-invoked Gateway to determine the current date-time from another Gateway without recourse to human intervention.

2. WB SATNET Message Dispatching

The Gateway's dispatcher packages up messages for output to the various networks supported by the Gateway. This dispatcher has been extensively modified to provide more optimal usage of network capacity, with special attention given to effective stream utilization on the WB SATNET.

The WB SATNET has a fairly small limit on the number of messages that can be transmitted on the channel each second. This severely limits the number of messages that may be specified for a stream when it is first created or later modified. This limitation, together with the long delay incurred in modifying a stream, makes the number of messages in a stream an expensive and inflexible resource that needs to be used efficiently. We must therefore transmit as few messages as possible while making the most use of each message that is transmitted. Towards this end, we made two extensions to the Gateway's dispatching for the WB SATNET: Multipurpose Packets (MPP) and Gateway-to-Gateway (GTG) IP fragments.

An MPP is a message that contains ST envelopes, IP messages, and GTG IP fragments packaged by a transmitting Gateway in such a way that it is recognizable as an MPP by a recipient IP/ST Gateway. The transmitting Gateway generates an MPP by aggregating messages until it exhausts the queue of messages waiting to be transmitted, the transmission capacity, or the maximum message size, whichever comes first. The receiving Gateway recognizes that the received message is an MPP, decomposes it into its parts, and continues processing the constituent messages as if they had come in separately. We are thereby able to transmit several logical messages intended for another Gateway in what the network regards as only one big message.

A GTG IP fragment is part of an IP message, possibly as small as one word, used to fill up remaining capacity in a stream after full ST envelopes and IP messages have been packaged in an MPP. We can thereby effectively use all the available stream capacity without any wastage. The transmitting Gateway generates a GTG IP fragment if an IP message is available for transmission. Such a fragment can be a 'middle' or 'last' fragment of a previously-started IP message or the 'first' fragment of a newly-started IP message. The receiving Gateway recognizes these fragments and performs the reassembly until all the fragments of the message have been received, at which point the IP message is processed as if it had been received completely in one message.

Major modifications to the dispatcher were needed to achieve MPPs and GTG fragments. Included in the new dispatcher is a look-ahead pass which tabulates the number of ST messages that are queued for transmission. Since ST messages have priority over IP messages for transmission, this tabulation determines how much of the stream capacity needs to be allocated to ST messages and therefore how much capacity remains for IP messages. During the second pass, the dispatcher builds MPP messages giving highest priority to the ST messages and filling out remaining space with IP messages or GTG fragments.

3. Non-Stream Network Message Dispatching

In the case of non-stream networks, one needs to balance delay in dispatching queued messages with network transmission capacity. Since available capacity at each of the Gateway's periodic wakeups is total capacity divided by the wakeup rate, we have the following tradeoffs. If one wakes up frequently to transmit messages, then the dispatch delay will tend to be low, but the capacity available at each wakeup will be small. On the other hand, if one wakes up infrequently, then the capacity at each wakeup will be high, but the dispatch delay will tend to be long. Therefore, a 'credit-debit' accounting system was incorporated into the dispatcher in order to dispatch outgoing messages optimally on non-stream networks. When messages are transmitted, capacity is debited by an appropriate amount. Correspondingly, the passage of time creates a credit for newly-available capacity. Using appropriate thresholds and limits, this approach permits the dispatcher to go into the 'red' and then work off the deficit while keeping the network as fully utilized as possible.

4. Probing of Other Hosts

A command is now available in the Gateway to enable a user to provide an internet address to be 'probed' via an ICMP ECHO message. (Receipt of an ECHO REPLY message from a probed host implies that the host is alive on the network and handles ICMP ECHO messages.) An optional IP source-route may be supplied in order to probe via a desired path, thereby determining whether all the intermediate points of the path are alive. By supplying a broadcast address as the internet address to be probed, one can probe a group of hosts simultaneously. These probes can be used at an IP/ST Gateway to determine the status of other Gateways and their attached networks.

5. Accommodation of 4 Networks

Extensions were made to enable the Gateway to support four attached networks. Such 'four-headed' Gateways have been installed at Lincoln and ISI, where the Gateways each support a WB SATNET, ARPANET, and two LEXNET ports.

6. Gateway Support

The 'todown' program provides communication and downloading capabilities from a host computer to a downline computer. This program has been installed at various sites and under various versions of the UNIX operating system (and one instance of the EPOS operating system) on PDP-11/44 computers.

During this reporting period, todown was 'ported' to two VAX computers and one SUN workstation, each running different versions of the UNIX operating system. These mark the first installations of todown on computers other than a PDP-11/44. Especially noteworthy in this successful porting is that no modifications were needed to the approximately 4000 lines of C code constituting todown in order to port the program to different hardware running under three different versions of UNIX.

D. SUN WORKSTATIONS

We are exploring the possibility of using the SPP to provide voice service as an added capability to workstations on a local computer network. Two SUN workstations running the UNIX operating system have been received and installed with an Ethernet link between the two stations. An RS-232 link to the Group's VAX computer has been provided to allow file transfers between the VAX and the SUN stations.

An SPP has been connected to one of the SUN stations and initial speech experiments have been run. So far we have not been able to transport full-duplex real time speech successfully from the SPP through the SUN. The principal bottleneck seems to be the slowness of the UNIX operating system in handling I/O through the RS-232 ports. Various means of increasing the transfer rate through these ports are being explored.

E. ETHERNET PVT

Work has begun on the development of an Ethernet interface for replacing the LEXNET interface within existing concentrator hardware and as an enhancement to existing or future Packet Voice Terminals. This new Ethernet Concentrator Interface (ECI) will include a 68000 microprocessor and will be packaged on two PVT-compatible wirewrap circuit cards. In addition to its basic LAN (Local Area Network) function, other optional ECI ports coupled with the computational power of a 16-bit microprocessor will permit the inclusion of Ethernet within LEXNET (Lincoln Experimental Packet Voice Network) or PLATFORM PVTs without increasing terminal card count.

The ECI central processor section is shown in Figure 1, and includes an 8-MHz CPU, a 4-channel DMA controller, 64 kbytes of EPROM, and 32 kbytes of SRAM storage. ECI peripheral options are shown in Figure 2. The multifunction interface includes an interrupt controller, programmable port bits, four timers, and full-duplex USART. For applications simultaneously requiring both asynchronous and synchronous serial capability, a multiprotocol communications controller may be added. Further flexibility can be achieved by including a byte-wide PVT DMA (Direct Memory Access) port (for protocol processor interfacing) and a GPIB (General Purpose Interface Bus) interface (as a vocoder bus). Through functional partitioning, the ECI central processor card is invariant, whereas the second peripheral card may be tailored in accordance with each application.

A paper design for the first prototype ECI has been completed, and all required non-Ethernet integrated circuits are on hand. A multifunction interface and a PVT DMA port will be included in the first version. The Ethernet port implementation will depend upon the availability of newly-developed VLSI chip sets, which should soon be obtainable on a sample basis.

A VALID Computer-Aided Engineering system is being used to provide accurate schematic and net list information for wirewrap fabrication and follow-on testing. The addition of a small number of needed IC types to the CAE (Computer Aided Engineering) component library will soon permit release of the ECI design for fabrication. Two test stands for physical support and powering of processor card pairs have been built.

The additional HP 64000 microprocessor development system workstation, configured expressly for 68000 hardware and software development, has been received and connected to the system.

Future plans call for ECI design validation and testing using two prototype processors operating in test stands. Diagnostic programs written in 68000 assembly language will be used for checkout of the ECI serial interfaces and the byte-wide PVT DMA port. Depending on component availability, the first Ethernet interface implementation may be based upon (1) Mostek/AMD hardware samples, (2) Fujitsu parts which are on hand, or (3) Mostek/AMD chip set emulator circuit cards. As part of the early Ethernet test effort, a minimal two-node network will be implemented. Interlan transceivers, cable, and terminators have been ordered for this purpose.

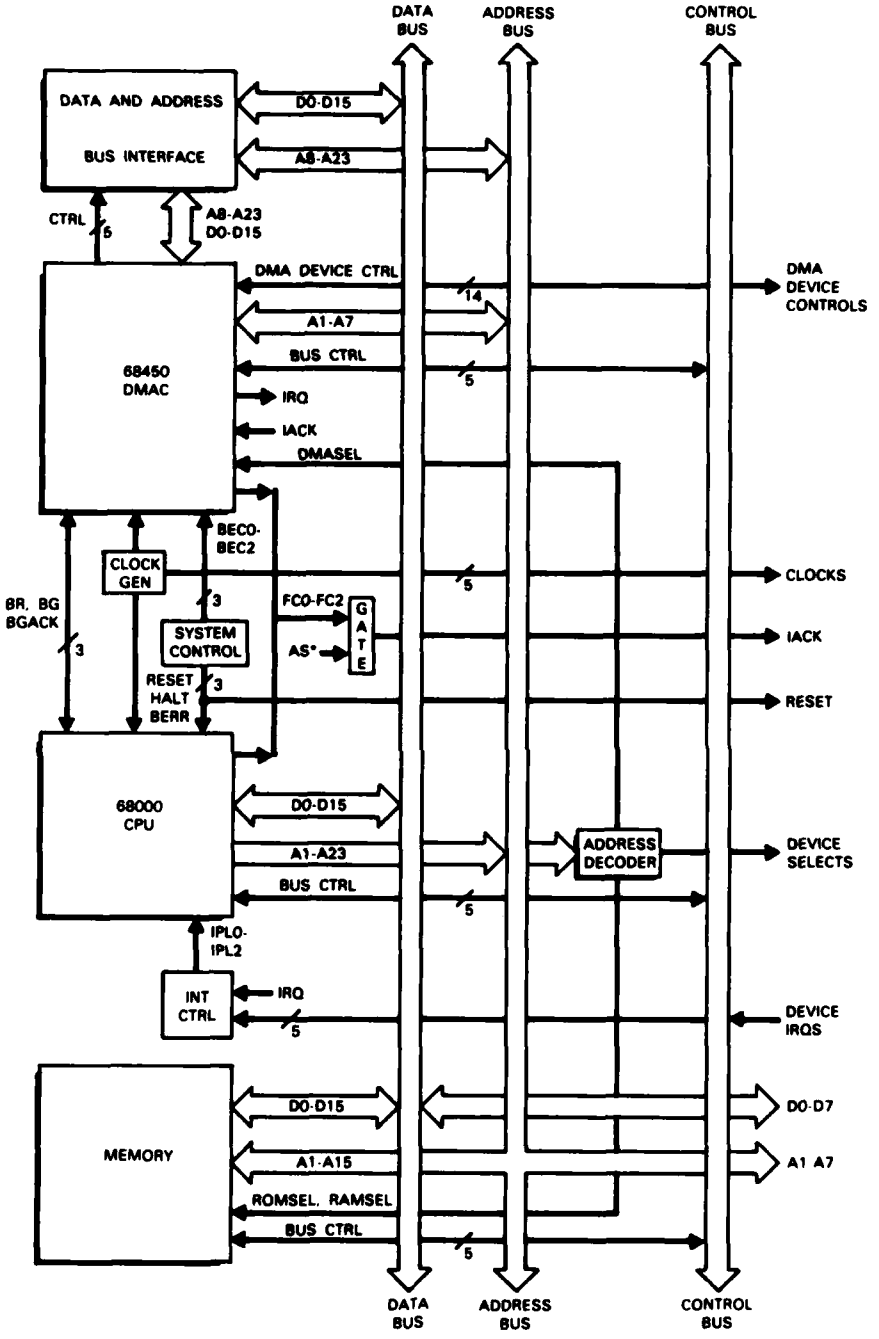

Figure 1. ECI Central Processor.

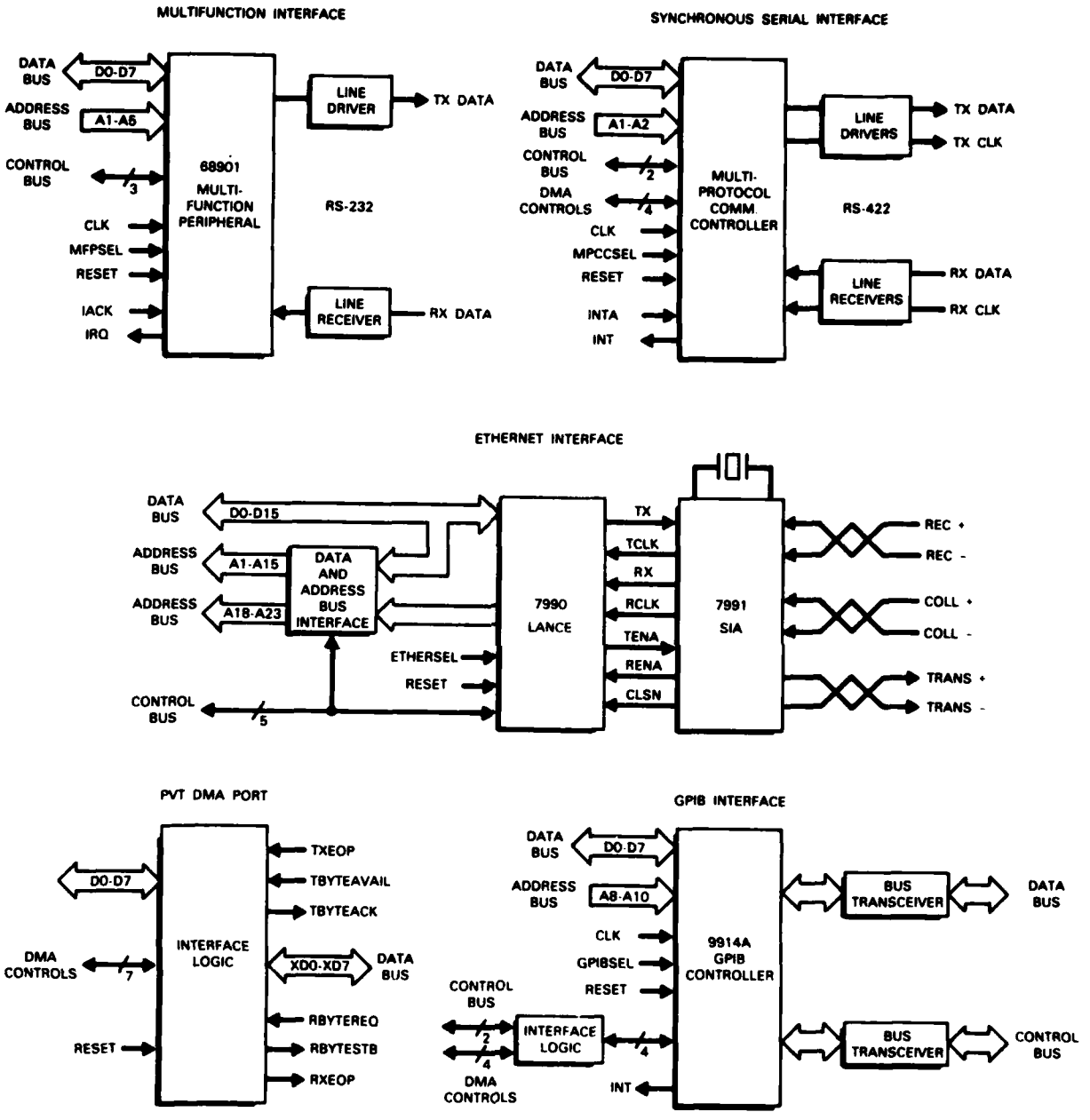

Figure 2. ECI Peripheral Options.

III. VOICE-CONTROLLED SYSTEMS

A. ROBUST SPEECH RECOGNITION STRUCTURE

We have designed several system structures to serve as a basis for our work in robust speech recognition. The baseline system structure is a noise-adaptive Dynamic-Time-Warping (DTW) template recognition system, which allows the basic template-matching approach to be augmented with a set of experimental algorithmic improvements including a perceptually-based distance metric, endpoint detection built into the DTW algorithm, automatic face-mask breath-noise rejection, and automatic template adaptation to the current noise environment.

As a basis for more advanced work, we have formulated two new system structures; these new structures potentially have greater extendability to talker-independence, larger vocabulary and connected speech. One structure consists of a Hidden Markov Model (HMM) isolated word recognition system with extensive enhancements that tailor the system to noisy environments. Proposed enhancements include automatic adaptation of word models to the current noise and stress environment, improved endpoint location built into the dynamic programming recognition algorithm, and improved perceptually-based distance metrics. A second system structure is designed for connected-word recognition. It consists of an efficient network-based recognition system (possibly HMM) followed by a feature-based system. The feature-based system would use discriminant analysis to improve the fine phonetic discrimination ability of the network-based recognition component.

The HMM system will serve as the basis for more advanced connected-word systems, and, in the isolated word application, has the potential of requiring much less training than DTW systems for large vocabularies due to the inclusion of phoneme and word models. An example of an HMM system designed for isolated words is presented in Figure 3.

The system in Figure 3 includes a spectrally-based front end, followed by a symbol classifier and a Viterbi Decoder, followed by a final classifier. The front end provides a detailed spectral estimate of the input every 10 msec. It also determines the word endpoints, estimates the noise background, estimates the probability of speech being voiced, and estimates the pitch of the input speech. The symbol classifier compares input spectra to stored symbol spectra and selects the one symbol whose spectrum best matches the input during the previous 10 msec. Sequences of symbols from the classifier are fed into the Viterbi decoder. This decoder determines the most probable path through each word model given the observed symbol sequence. It also calculates the probability of this path and feeds these probabilities to the final classifier. The final classifier selects a word hypothesis based on word probabilities and known probabilities of word sequences. Word models are made up of phoneme models and require training to specify: the topology of word and phoneme networks, transition probabilities between nodes, definitions of symbol spectra, and symbol probabilites for all transitions. Word models are updated during use whenever the probability that the system selected the correct word model (as estimated by word probabilities) is high. A major new feature of the system shown in Figure 3 is that word models created in

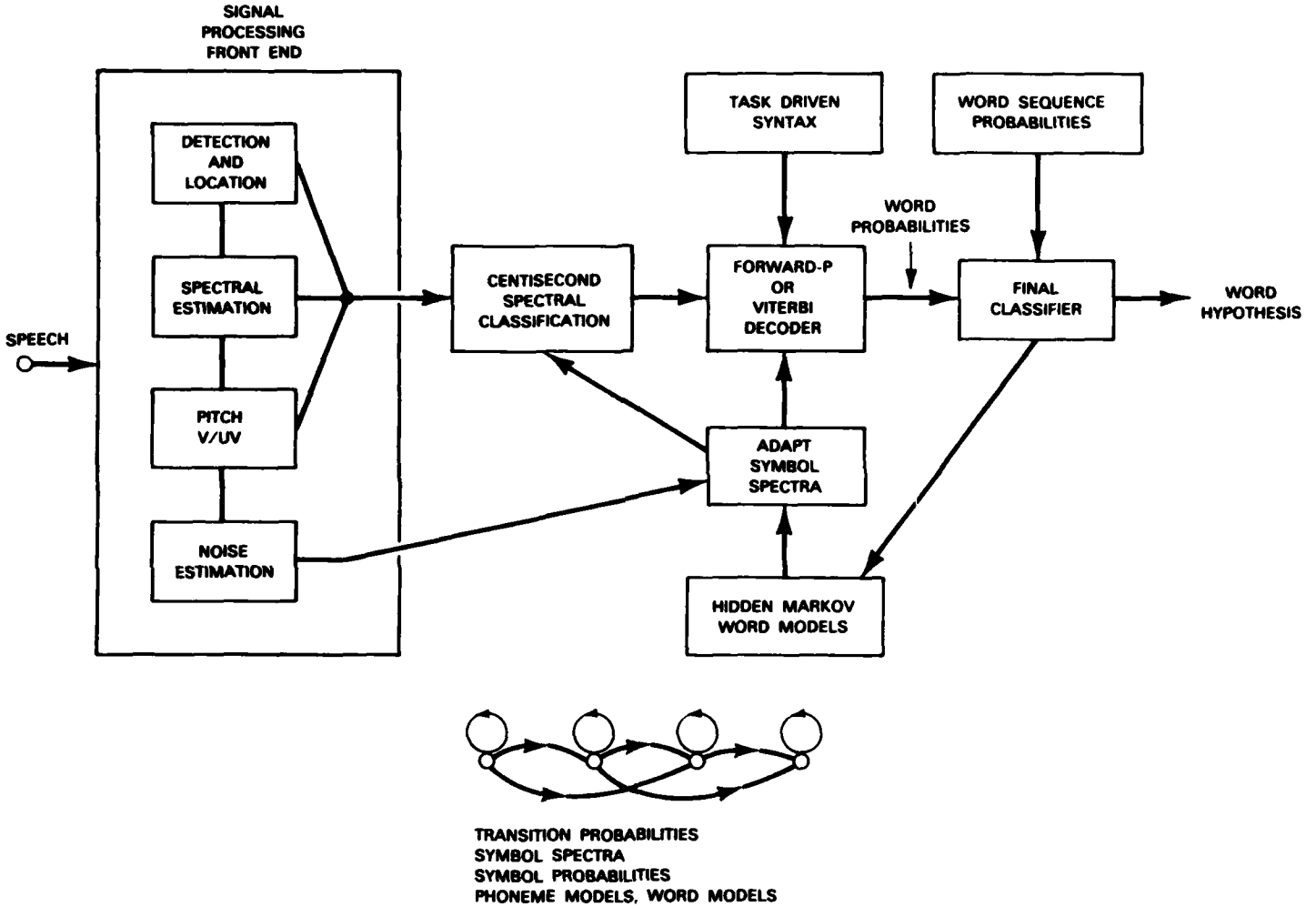

Figure 3. Robust Hidden Markov Model (HMM) Recognition System.

135654-N

quiet are adapted for the noise environment. This is primarily done by estimating the current noise background and spectrally adding the current noise spectrum to symbol spectra to produce new symbol spectra appropriate for the given environment. Such an approach is suggested by the need to test and train template systems in the same environment for good performance and by recent statistical maximum likelihood approaches we have developed. The form of word models is also changed under noisy conditions by collapsing symbols that have similar spectra in noise. Other modifications to improve the performance of the system in noise are: (1) endpoint detection is incorporated in the classification search by allowing noise symbols at the ends of word models, (2) symbol classification is performed using a perceptual distance metric that adapts to the current S/N ratio, and (3) the Viterbi search probability calculation is modified to weight temporal regions with greatest S/N ratio more heavily and also to weight regions where the input spectrum is changing rapidly more heavily.

Our second system structure, which is designed for connected-word recognition and includes both a network-based and a feature-based component, is shown in Figure 4. The system in Figure 4 includes the same front end shown in Figure 3, followed by a robust network-based recognition system and a feature-based discriminant analysis system. Recent results obtained at Carnegie-Mellon University*, using a feature-based approach to distinguish among a small set of highly confusable words, are promising. The feature-based system will measure perceptually important acoustic features (spectral energy peaks, silent intervals, center frequency of energy of bursts, segment durations. . .) and use these measurements in combination with discriminant analysis to select one of a small set of candidates as the spoken word. The lower feedback paths in Figure 4 indicate that discriminant statistics and HMM probabilities are automatically tuned during use when the selected word sequence hypothesis is highly likely.

The system in Figure 4 uses a feature-based system as a word verifier that examines only those words with a high likelihood of being recognized incorrectly by the network-based recognition system. It compares each such word to a small set of other candidate words selected during the network search and from confusion statistics obtained previously. Comparisons use segmentation information obtained using Viterbi decoding and prior discriminant statistics obtained during training. Advantages of this approach are that it provides independent word verification, it uses the network-based system to severely restrict the computation required by the feature-based system, and it uses an overall system with components that are all automatically trained.

B. SPEECH AND NOISE MODELING FOR ROBUST RECOGNITION

In the Hidden Markov modeling approach to speech recognition, the efficacy of statistical pattern-matching techniques for the word classification problem has been demonstrated

*R.M. Stern and M.J. Lasry, "Dynamic Speaker Adaptation for Isolated Letter Recognition Using Map Estimation," Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, Boston, 14-16 April 1963.

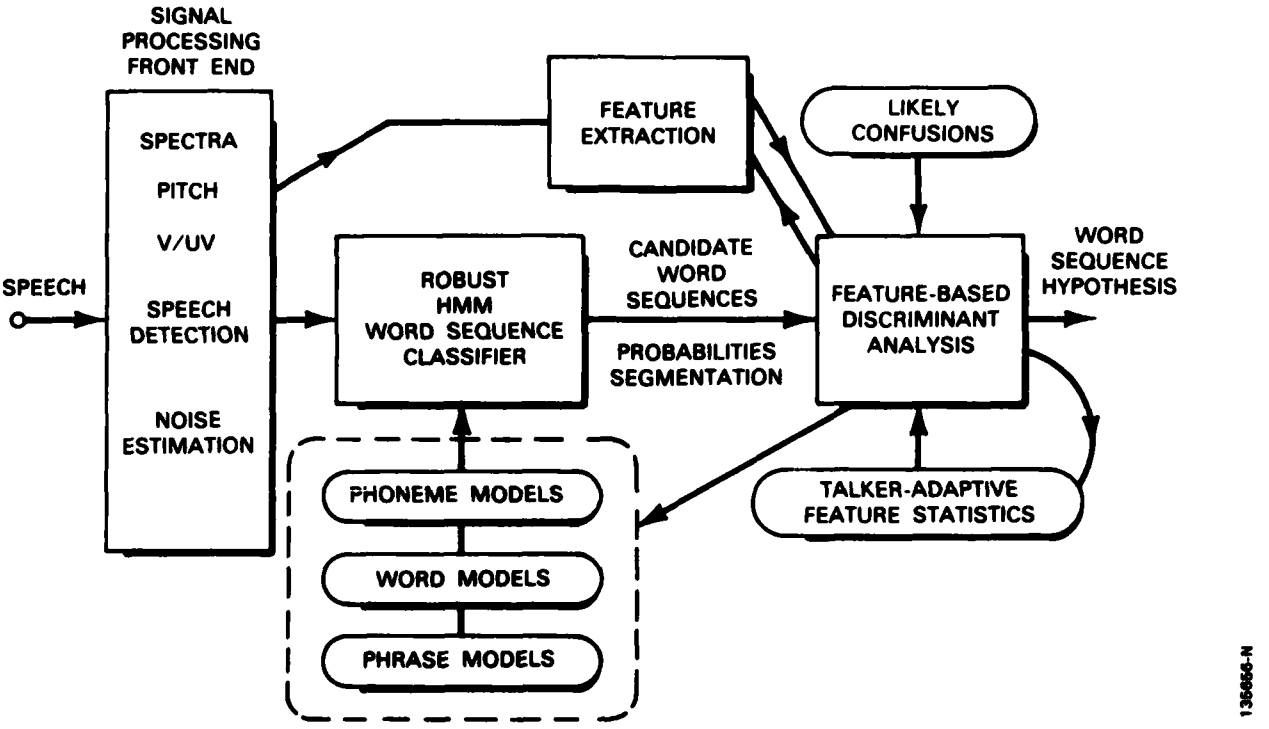

Figure 4. Two-Stage HMM-Discriminant Speech Recognition System.

by IBM and others. When the speech is degraded by environmental acoustic noise, one would expect the statistical approach to be particularly effective in developing a suitable classification criterion. We have initiated a study to model the word generation and noisy measurement processes from a statistical point of view and have developed new spectral matching criteria that reflect the effects of the additive noise. The results suggest that the speech should be trained in the clear and that the noise power spectrum should be estimated on-line and added to the speech templates just prior to classification. An experimental program is proposed for evaluating these ideas.

We have chosen to characterize the speech generation process by the speech production model which is specified in terms of a data structure consisting of a voiced/unvoiced decision, a pitch if voiced, and a spectral template that determines the shape of the vocal tract (i.e., for an acoustic tube synthesizer). Speech is produced by specifying the time-sequence of data structures. This process is then characterized by a left-to-right Markov model whose output is a pointer to a particular data structure. The evolution of the Markov states and the mapping to the speech states is probabilistic.

The speech output of this production model is then corrupted by additive acoustic noise of some spectral shape that is continually being updated by the system. The measurement system consists of a high-resolution FFT and a pitch-directed peak-picking algorithm that effectively isolates each of the sinusoidal components of the speech waveform. This has the effect of enhancing the measurement signal-to-noise ratio. Since there is no information in the phase of these peak measurements, only the envelope is passed to the classification algorithm. This measurement model can be analyzed statistically to characterize the mapping from the speech data structure to the noisy observations. It is this step that gives rise to the spectral matching criterion that includes the effects of the noise. By modeling the speech and noise as Gaussian processes, it has been shown that the distance function is the Itakura-Saito spectral criterion in which the noise power spectrum is added algebraically to the speech spectral template. In the absence of noise, therefore, the analysis can be shown to reduce to a frequency domain version of Dynamic Time Warping.

An experimental program is being formulated to evaluate some of the ideas suggested by the analytical studies. In particular, an existing LPC DTW recognition system can be modified to use the frequency-based spectral matching criterion. Then the recognition in noise can be tested by adding the average noise power spectrum to each of the speech templates before matching. The performance of this system can then be compared with one that trains on the basis of noisy speech.

C. ADVANCED SPEECH RESOURCE UNIT SIMULATION FACILITY

Lincoln's DARPA-sponsored Restructurable VLSI Program includes the development of a wafer-scale Dynamic Time Warping (DTW) device for isolated-word and connected-word speech recognition. To operate as a speech-recognition system, the DTW device must be supported by appropriate data-processing and voice-processing subsystems. We refer to the composite system as an Advanced Speech Resource Unit (ASRU), which we expect will be composed of a single board including DTW template matching, voice processing, and data processing.

As part of the current Wideband Program, we are implementing software versions of the data-processing subsystem to provide test support for the DTW as well as a basis for hardware design of the data-processing portion of the ASRU. The voice processing is being provided using our compact LPC hardware. The data-processing software, which initially has been implemented in a PDP-11, has as its primary component the level-building algorithm needed to perform connected-word recognition. This algorithm has been successfully implemented and tested, along with an LDSP simulation of the DTW wafer. The system has operated successfully for digit strings up to nine in length.

In order to examine in detail a number of design tradeoffs in the DTW system (e.g., distance metric, internal word length, search path restrictions), we currently are implementing a more flexible simulation facility in our VAX computer. This facility will include level-building as part of the VAX software. Much of the software for the VAX facility is currently operating, and we expect to begin detailed investigations of DTW design tradeoffs

using standard speech data bases during the next quarter. These investigations will impact the projected data-processing requirements for ASRU, as well as the detailed wafer design.

D. LDSP-BASED NOISY SPEECH RECOGNITION FACILITY

An LDSP-based speech-recognition facility developed earlier* for testing the effect of noise preprocessing on recognition of noisy speech has been brought back into operation and modified for the purpose of recognition tests using new noisy-speech data. Initial tests with the Advanced Fighter Technology Integrator (AFTI) F-16 data base have indicated that the simple endpoint detection algorithm used in the earlier work is clearly inadequate to handle the new data. The major problem is that the utterances are generally preceded and followed by strong breath noise which spoofs the endpoint detector. We currently are implementing a variable endpoint DTW algorithm as an initial candidate to address this problem. This algorithm will allow endpoint selection to take place along with the DTW match, instead of prior to the DTW match. This algorithm will initially be implemented and tested in the C language on the VAX to minimize programming time and enhance flexibility.

*G. Neben, "The Performance of an Isolated Word Recognition System Using Noisy Speech," Technical Report 647, Lincoln Laboratory, M.I.T. (April 1983) AD-A128983/4.

GLOSSARY

AFTI	Advanced Fighter Technology Integrator
AR	Adams-Russell Company
ARPANET	ARPA Network
ASRU	Advanced Speech Resource Unit
BBN	Bolt, Beranek and Newman
BSAT	Butterfly Satellite Interface Message
CAE	Computer Aided Engineering
CMU	Carnegie-Mellon University
DMA	Direct Memory Access
DRT	Diagnostic Rhyme Test
DTW	Dynamic Time Warping
ECI	Ethernet Concentrator Interface
ESI	Earth Station Interface
GPIB	General Purpose Interface Bus
GTG	Gateway-to-Gateway
HMM	Hidden Markov Model
ICMP	Internet Control Message Protocol
IP	Internet Protocol
ISI	Information Sciences Institute
LAN	Local Area Network
LDSP	Lincoln Digital Signal Processor
LEXNET	Lincoln Experimental Packet Voice Network
LPC	Linear Predictive Coding
MPP	Multipurpose Packets
PC	Printed Circuit
PCM	Pulse Code Modulation
PSAT	Packet Satellite Interface Message Processor
PVT	Packet Voice Terminal
SPP	Speech-Processing Peripheral
SRI	SRI International
ST	Stream Protocol
WB SATNET	Wideband Satellite Network

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ESD-TR-84-018	2. GOVT ACCESSION NO. AD-A146 054	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Wideband Integrated Voice/Data Technology		5. TYPE OF REPORT & PERIOD COVERED Semiannual Technical Summary 1 October 1983 — 31 March 1984 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* Clifford J. Weinstein		8. CONTRACT OR GRANT NUMBER*(s)* F19628-80-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173-0073		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element No. 62708E Project No. 3T10 ARPA Order 3673
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22209		12. REPORT DATE 31 March 1984 13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* Electronic Systems Division Hanscom AFB, MA 01731		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

packet speech	internetwork	speech recognition
Linear Predictive Coding	protocol	voice conferencing
LEXNET	gateway	packet voice terminal
Wideband SATNET	ARPANET	

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This report describes work performed on the Wideband Integrated Voice/Data Technology Program sponsored by the Information Processing Techniques Office of the Defense Advanced Research Projects Agency during the period 1 October 1983 through 31 March 1984.

DD FORM 1473 1 Jan 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*